FEBS Letters 582 (2008) 792–798

Putative alternative *trans*-splicing of leukocyte adhesion-GPCR pre-mRNAs generates functional chimeric receptors

Pei-Ling Chiu[a,1], Boon Han Ng[b], Gin-Wen Chang[c], Siamon Gordon[c], Hsi-Hsien Lin[b,*,1]

[a] *Department of Medical Biotechnology and Laboratory Science, College of Medicine, Chang Gung University, 259 Wen-Hwa 1st Road, Kwei-San, Tao-Yuan, Taiwan*

[b] *Department of Microbiology and Immunology, College of Medicine, Chang Gung University, 259 Wen-Hwa 1st Road, Kwei-San, Tao-Yuan, Taiwan*

[c] *Sir William Dunn School of Pathology, The University of Oxford, South Parks Road, Oxford OX1 3RE, UK*

Received 31 October 2007; revised 7 January 2008; accepted 1 February 2008

Available online 11 February 2008

Edited by Ulrike Kutay

Abstract The EGF-TM7 receptors, a subfamily of adhesion-GPCRs mostly restricted to leukocytes, are known to express multiple functional protein isoforms through extensive alternative *cis*-splicing. Here, we demonstrate that EGF-TM7 pre-mRNAs also undergo the rare *trans*-splicing, leading to the generation of functional chimeric receptors. RT-PCR and in silico analyses of EMR2 transcripts identified unique fragments containing the EGF-like motif 3 of a closely related EGF-TM7 gene, CD97, in addition to the alternative *cis*-spliced products. The sequence swapping is restricted to the EGF-3 exon, generating unique EMR2(1-2-3*-5) and EMR2(1-2-3*-4-5) molecules, which are functional in ligand-binding as the wild-type EMR2(1-2-3-4-5) and CD97(1-2-3-4-5) receptors. Our results suggest that human leukocytes employ *trans*-splicing as well as *cis*-splicing to increase the repertoire of functional adhesion-GPCRs.

Keywords: Adhesion-GPCR; EGF-TM7; LNB-TM7; Alternative splicing; *trans*-Splicing

1. Introduction

The completion of the human genome project and others has made it evident that the number of genes in a multi-cellular organism's genome is far less than previously believed [1,2]. The diversity of the proteome however can be greatly enhanced from a fixed-size genome through various co-/post-transcriptional and post-translational modifications. Alternative splicing selectively combines different exons of a pre-mRNA to form different mRNA products, which in turn encode protein isoforms of similar or distinct functions. Thus, alternative splicing is thought to be one of the most important co-/post-transcriptional mechanisms for increasing protein-coding capacity of the genomes [3–5]. In fact, it has been estimated that ~35–60% of all human genes undergo alternative splicing [6,7].

Although alternative splicing takes place predominantly in *cis*-mode within an individual pre-mRNA, recent studies have shown that the same splicing reaction can also occur in *trans*-mode between separate pre-mRNAs of either the same gene (intragenic) or different genes (intergenic) [5,8]. Thus, intragenic *trans*-splicing could produce mRNAs containing duplicated exons of one pre-mRNA, whereas intergenic *trans*-splicing would be able to join exons of two different pre-mRNAs to form chimeric mRNAs.

Alternative *trans*-splicing has been identified in species ranging from unicellular organisms such as trypanosomes to more complex ones such as *Drosophila* and mammals [5,8]. In trypanosomes and *Caenorhabditis elegans*, *trans*-splicing is frequent, but usually involves the joining of a common 5′ non-coding spliced leader to the downstream polycistronic genes [9,10]. No protein diversity is generated in this type of *trans*-splicing. In *Drosophila*, *trans*-splicing is essential for the functional diversities of *mod*(*mdg4*) and *lola* genes that are involved in the establishment/maintenance of chromatin structure and axon guidance pathways, respectively [11–14]. In mammals, *trans*-splicing is detected much less frequently. In addition, it is noted that all mammalian genes known to undergo *trans*-splicing also actively engage in *cis*-splicing [5,8]. Indeed, it was proposed that mammalian *trans*-splicing might be a result of splicing noise. Nevertheless, *trans*-splicing was clearly demonstrated in both in vitro and in vivo systems [15–17]. Mammalian *trans*-splicing has been identified in virus-infected cells, which produced combined viral and cellular RNAs, as well as in several endogenous cellular mRNAs with duplicated exons [18–22]. Examples of intergenic *trans*-splicing between different genes have also been reported elsewhere [23–26]. Recently, RNA *trans*-splicing has even been applied as a novel RNA-based gene therapeutic approach to correct mutant/diseased genes [27,28].

The EGF-TM7 receptors are leukocyte-restricted 7-span transmembrane (TM7) proteins with multiple extracellular epidermal growth factor (EGF)-like motifs [29,30]. They belong to the class of class B-related TM7 receptors with a long N-terminal extracellular domain (LNB-TM7)

*Corresponding author. Fax: +886 32118700.
E-mail address: hhlin@mail.cgu.edu.tw (H.-H. Lin).

[1]These authors contributed equally to this work.

Abbreviations: CS, chondroitin sulfate; EGF, epidermal growth factor; EMR, EGF-like module containing mucin-like hormone receptor; ERS, exon-recognition sequence; EST, expressed sequence tag; Fc, *fragment crystallisable*; GAG, glycosaminoglycans; GPCR, G protein-coupled receptor; LNB-TM7, class B-related TM7 receptors with a long N-terminal extracellular domain; TM7, seven-span transmembrane

0014-5793/$34.00
doi:10.1016/j.febslet.2008.02.004

receptors/adhesion-GPCRs, which are defined as class B GPCR-like molecules having a long extracellular domain with protein motifs involved in cellular adhesion [31,32]. In the human genome, a total of 5 EGF-TM7 genes, namely EMR1, 2, 3, 4, and CD97, are clustered within two regions on chromosome 19p13 [33]. It is thought that the ancestor EGF-TM7 gene was created by exon shuffling linking the EGF-like motifs to a secretin receptor-like 7TM gene [33,34]. From the phylogenetic analysis and the EGF-TM7 gene arrangement on chr.19, it was further suggested that the EGF-TM7 family was derived from an ancestor gene through gene duplication and conversion [35]. Thus, EMR2 was shaped by concerted evolution to be a chimeric molecule that shares strong homology in the EGF-like motifs and the 7TM region with CD97 and EMR3, respectively [35].

The EGF-TM7 receptors are characterized by extensive alternative *cis*-splicing, resulting in protein isoforms with different numbers of the EGF-like motifs [30,32]. Of note, the closely related EMR2 and CD97 genes both contain 5 highly homologous EGF-like motifs that are 97.5% identical [36]. Both genes have been shown to generate mRNA species with 1-2-5, 1-2-3-5, and 1-2-3-4-5 arrangement of the EGF-like motifs. Another EMR2 isoform was predicted to contain EGF-like motifs 1-2 only [36–38]. Most importantly, the resulting protein variants were shown to possess different ligand-binding activities. Thus, the CD55-binding affinity of CD97 isoforms was found to be in the order of CD97(1-2-5) > CD97(1-2-3-5) > CD97(1-2-3-4-5), while EMR2(1-2-5) showed little if any CD55-binding [36,39–41]. In contrast, both EMR2(1-2-3-4-5) and CD97(1-2-3-4-5), but not other isoforms, displayed a similar interaction with another cellular ligand, chondroitin sulfate (CS)-glycosaminoglycans (GAGs) [42,43]. Alternative splicing therefore provides a means for the EGF-TM7 receptors to expand their ligand repertoire and presumably their cellular functions.

In the present study, we provide evidence for alternative *trans*-splicing between EMR2 and CD97. We find that alternative *trans*-splicing joined specifically the CD97(EGF-3) exon between the EGF-2 and EGF-4/EGF-5 exons of EMR2, creating unique EMR2/CD97(EGF-3) chimeric molecules, namely EMR2(1-2-3*-5) and EMR2(1-2-3*-4-5) (* denotes the sequence of CD97 EGF-like motif). We also show that the chimeric EMR2(1-2-3*-4-5), but not EMR2(1-2-3*-5), shares similar CS-binding activity with the wild-type EMR2(1-2-3-4-5) and CD97(1-2-3-4-5). Our data suggest that human leukocytes use alternative *trans*-splicing, in addition to alternative *cis*-splicing, to further increase the diversity of the EGF-TM7 receptors.

2. Materials and methods

2.1. Reagents and cell culture

General chemicals of analytical grade were obtained from Sigma–Aldrich (St. Louis, MO, USA), unless otherwise stated. All culture media were from Invitrogen and were supplemented with 10% heat inactivated fetal calf serum (FCS), 2 mM L-glutamine, 50 IU/ml penicillin and 50 μg/ml streptomycin. HEK-293T cells, wild-type CHO-K1 and its mutant derivatives (PgsB-618 and PgsD-677) were cultured in DMEM and Ham's F12 medium, respectively. Total RNAs from pooled human peripheral leukocytes were purchased from BD Biosciences-Clontech.

2.2. RT-PCR and DNA analysis

To identify specific EMR2 transcripts, RT-PCR was performed on pooled human peripheral leukocytes total RNAs using nested EMR2-specific primer sets. In brief, 1 μg of DNA-free total RNAs were reverse transcribed using random hexamers and MMLV reverse transcriptase as described previously [36]. The cDNA products were then amplified by PCR using the 1st EMR2 primer set: EMR2/5′-1,5′-CTGCTCCTGCCGGCAGCTCAGCTGGAACCAT-3′ and EMR2/3′-1,5′-ACGGGATCCTCCTGCACATCGTAGTGGGCCATGA-3′. The resulting RT-PCR products were diluted 200-fold and re-amplified again using the 2nd EMR2 primer set: EMR2/5′-2, 5′-AACCATGGGAGGCCGCGTCTTTCTCGTCTTTCTCGCA-3′ and EMR2/3′-2, 5′-TAGGCCATCCAGCAGGTGACTGGCACACA-3′. The final products were separated by electrophoresis in 1% agarose gels. All amplified DNA fragments were excised from the gel together in a single gel slice, extracted, and subcloned into the pCR4-TOPO vector (Invitrogen). A total of 70 random individual colonies were selected and sequenced. For Southern blot analysis, PCR products were digested with or without BamHI prior to gel electrophoresis. DNA fragments were blotted onto Duralon-UV filters (Stratagene) and hybridized with a EMR2 specific probe, which was PCR-amplified by 5′-CCTGGGGACCTGGAGACCCTG-3′ and 5′-CAAGGTGACACTCCTGTCTACTT-3′, representing a DNA fragment at the stalk region. For the search in database of *trans*-spliced EMR2 transcripts, the BLAST program was performed using the EMR2(1-2-3*-4-5) cDNA sequence as a template.

2.3. Construction of mFc-fusion expression vectors

EMR2(1-2-3-4-5)-mFc and CD97(1-2-3-4-5)-mFc expression constructs have been described previously [43]. The EMR2(1-2-3*-4-5)- and EMR2(1-2-3*-5)-mFc constructs were generated similarly by cloning the appropriate PCR-amplified EGF-like motifs in frame upstream of the mFc fragment via HindIII and EcoRI sites. The constructs were expected to express mFc-fusion proteins containing the indicated EGF-like motifs and a short stalk region.

2.4. Production of mFc-fusion protein probes

mFc fusion proteins were produced and purified as described previously [43]. In brief, 60–70% confluent HEK-293T cells were transfected with the mFc-fusion expression constructs by calcium phosphate precipitation method. Typically, ~40 μg of plasmid DNA was used for one T-175 flask of HEK-293T cells. Transfected cells were washed and fed with fresh OPTI-MEM I medium ~16–24 h post-transfection. Serum-free conditioned media containing the secreted fusion proteins were collected after 5–6 days of culture, filtered, and passed through the Protein A-sepharose affinity chromatography column according the manufacturer's protocol (GE Healthcare). After extensive washes, mFc fusion proteins were eluted from the column with 100 mM glycine buffer, pH 2.9, and immediately neutralized with 1.0 M Tris–HCl, pH 9.0 buffer. Purified proteins were dialyzed in 50 mM Tris–HCl, pH 7.4 buffer containing 10 mM $CaCl_2$ and 150 mM NaCl at 4 °C. The final protein concentration was measured by the BCA assay (Pierce, Rockford, USA) according to the manufacturer's protocol. Purified proteins were confirmed by Western blotting, aliquoted, and stored at −80 °C until use.

2.5. Cell-capture assay

To examine the CS-binding ability of the EMR2/CD97 isoforms, a cell-capture binding assay was performed as described previously [43,44]. Briefly, the purified mFc-fusion proteins (5 μg) were coated onto the pre-washed protein G-conjugated para-magnetic beads (20 μl) (Protein G Dynabeads®, Dynal A.S., Oslo, Norway). The protein–bead solution was suspended in a final volume of 200 μl PBS/0.1% BSA and mixed thoroughly at room temperature for 1 h. Following three washes in PBS/0.1% BSA to remove unbound proteins, the protein–bead complexes were added to wild-type or mutant CHO-K1 cell lines in a total volume of 0.5 ml at the density of 1×10^4 cells/μl. The protein–bead-cell mixtures were mixed thoroughly by end-over-end rotation at 4 °C for 1 h. After extensive washes and magnetic separation, the captured cells were released from the beads with PBS/10 mM EDTA at room temperature for 10 min and magnetic separation. Cells were then stained with trypan blue solution and viable cell numbers were determined by haemocytometer counting.

3. Results and discussion

3.1. Analysis of EMR2 alternative splicing and identification of trans-spliced EMR2 transcripts

In a previous study examining EMR2 alternative splicing, we have identified multiple alternatively *cis*-spliced transcripts encoding protein isoforms with EGF-like motifs 1-2, 1-2-5, 1-2-3-5, and 1-2-3-4-5 configurations [36]. Interestingly, analysis of certain transcripts has unexpectedly yielded non-identical matches to the wild-type EMR2 cDNA sequences. A more detailed examination showed that the mismatched region is localized at the exon encoding the EGF-like motif 3, and when analyzed further is found to be identical to the highly homologous 3rd EGF-like motif of CD97 (data not shown).

To confirm the previous observation, we performed a stringent RT-PCR analysis using nested EMR2-specific primers on RNA samples isolated from pooled human leukocytes. As EMR2 and CD97 are highly homologous, we used nested EMR2-specific primer sets to minimize the possibility of non-specific amplification. The primers used here encompass a region containing the signal peptide, the entire 5 EGF-like motifs, and a partial stalk region that are unique to EMR2 (see Section 2). As expected, multiple EMR2-specific cDNA fragments were identified (Fig. 1A, lane 1). As controls for primer specificity, plasmids containing the full-length EMR2(1-2-3-4-5) and CD97(1-2-3-4-5) cDNAs were used independently as templates. As shown in lanes 3 and 4 of Fig. 1A, a specific PCR product was detected only in the reaction containing EMR2, but not CD97 cDNA template. The multiple specific EMR2 cDNA fragments were excised from gels, subcloned, and a total of 70 independent colonies were sequenced. As expected, the majority of clones had sequences perfectly matched to the distinct *cis*-spliced EMR2 cDNAs (Fig. 1B). However, a total of 15 out of the 70 clones were found to contain the sequence of the 3rd EGF-like motif of CD97 (EGF-3*). Specifically, EMR2(1-2-3*-5) and EMR2(1-2-3*-4-5) sequences were identified in 11 and 4 clones, respectively (Fig. 1B).

The fact that each EGF-like motif of EGF-TM7 proteins is encoded by an independent single exon [34,36], and that only the 3rd EGF-like motif but not others were "swapped" between EMR2 and CD97 argue strongly that the RT-PCR results are genuine. The 7-nucleotide difference between the EGF-3 motifs of EMR2 and CD97 further suggests that these chimeric sequences are not the results of PCR artefact (Fig. 2A). Finally, as the orientations of EMR2 and CD97 genes on chr.19p13 are in opposite directions, it is unlikely that the chimeric transcripts are the products of transcription-induced gene fusion found in some adjacent genes [45,46]. Taken together, we conclude that these chimeric transcripts are most likely derived from intergenic alternative *trans*-splicing between EMR2 and CD97.

To provide further evidence for the presence of the *trans*-spliced transcripts, we took advantage of the BamHI site

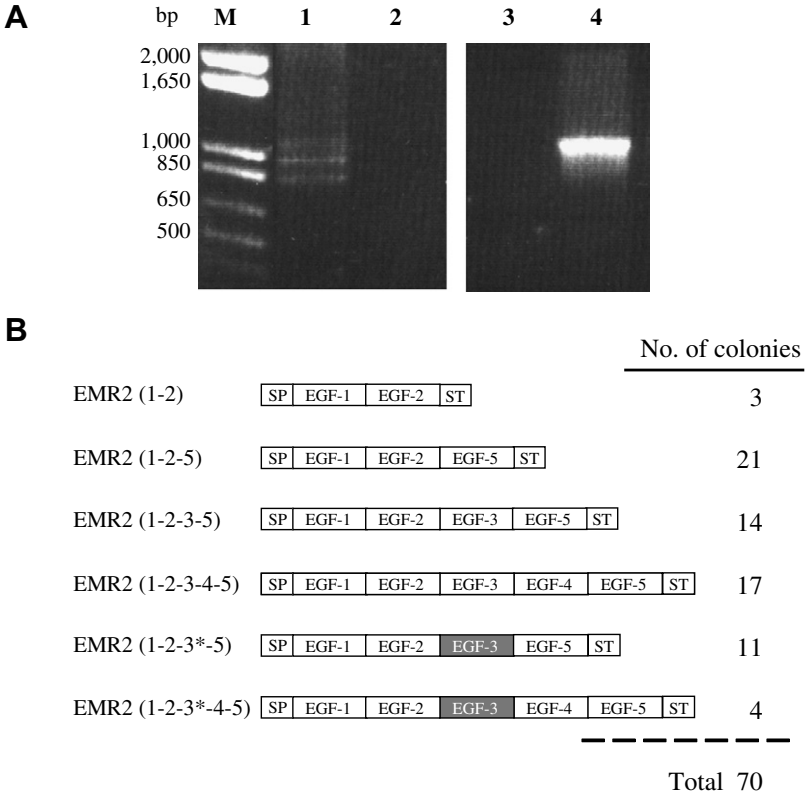

Fig. 1. RT-PCR analysis of EMR2 alternative splicing. (A) RT-PCR products amplified from RNA of pooled human leukocytes were separated in a 1% agarose gel. Lanes 1 and 2 show the results of RT(+) and RT(−) templates, respectively. At least three distinct bands ranging from ~800 to 1100 bp were detected in the RT(+) sample. The specificity of the primers used in the reaction was confirmed in lanes 3 and 4 using the full-length CD97 and EMR2 cDNA, respectively as templates. (B) The sequence analysis of RT-PCR products as seen in (A), lane 1. SP and ST stand for signal peptide and stalk region, respectively. The CD97 EGF-3 motif is shown as a grey box while EMR2 exons are shown as white boxes. M represents the DNA marker.

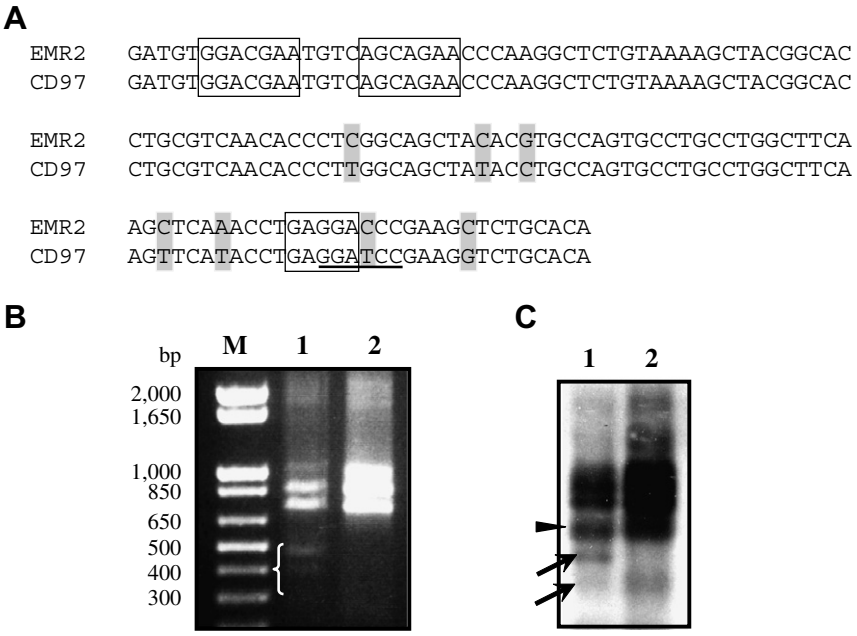

Fig. 2. Confirmation of *trans*-spliced EMR2 transcripts. (A) Sequence comparison of the EGF-3 motifs of EMR2 and CD97. The seven different nucleotides between EMR2 and CD97 EGF-3 motifs are highlighted in grey background. The unique BamHI site in CD97 is underlined. The three potential ERS motifs are indicated by boxes. (B) The RT-PCR products were digested with (lane 1) or without (lane 2) BamHI before gel electrophoresis. The gel was over-exposed to show the BamHI-digested bands (indicated within a white curly bracket). (C) The same gel was then analyzed by Southern blotting using a EMR2-specific probe. The arrows represent BamHI-digested bands that reacted positively to the probe, while the arrowhead shows non-specific bands reacting to the probe. M represents the DNA marker.

found only in the CD97 but not EMR2 EGF-3 motif (Fig. 2A). RT-PCR products were digested with or without BamHI, separated by electrophoresis, and hybridized with a EMR2-specific probe. Several additional faint but distinct bands were clearly visible after BamHI digestion (Fig. 2B). The sizes of the extra bands (~380 bp, ~430 bp, and ~530 bp) correlated nicely to those of the expected BamHI-digested fragments. Furthermore, two of the bands reacted specifically to the unique EMR2 3′-end probe in Southern blotting analysis (Fig. 2C), indicating an authentic EMR2 sequence. These results clearly show that there are minor populations of EMR2-specific cDNA fragments possessing a unique BamHI site, which is foreign to the wild-type EMR2 sequence. Taking into the consideration of the previous findings, it is believed that these are the *trans*-spliced EMR2 transcripts containing the CD97 EGF-3* motif.

3.2. In silico analysis of trans-spliced EMR2 transcripts

A survey of human expressed sequence tag (EST) databases has also identified examples supporting the presence of *trans*-spliced EMR2 transcripts. While 25 EST clones were found to have perfect sequence matches to various wild-type EMR2 *cis*-spliced transcripts, a total of 5 EST clones were found to match, or partly match to the *trans*-spliced EMR2 transcripts, but not to the wild-type EMR2 or CD97 sequences (Table 1). All 5 contain the EGF(1-2-3*) sequence and one especially represents a full-length EMR2 cDNA with EGF(1-2-3*-4-5)-TM7 configuration (clone BC082991). As these EST clones are of high quality cDNA sequences derived from different tissues by independent groups, it is reasonable to believe they are the true representatives of transcripts expressed by these tissues.

With the 3 different lines of evidence described above, we suggest that alternative *trans*-splicing events indeed occur between EMR2 and CD97 pre-mRNAs and potentially lead to the production of EMR2/CD97 chimeric proteins.

Table 1
Human EST clones containing *trans*-spliced EMR2 sequences

Trans-spliced EGF-like motifs	Accession number	RNA source (mRNA length, bp)
1-2-3[a]	DR001061	Human fetal brain (653)
1-2-3[a]-4	BU507349	Leiomyosarcoma (912)
	DA402257	Human thalamus (548)
	DA676179	Neutrophils (593)
1-2-3[a]-4-5	BC082991	Human uterus (3467)

[a]Represents the CD97 EGF-like motif.

3.3. Functional analysis of the trans-spliced EMR2 transcript products

We have previously demonstrated that both EMR2(1-2-3-4-5) and CD97(1-2-3-4-5) interact specifically with CS-GAG via the 4th EGF-like motif, whereas other isoforms fail to do so [42,43]. It is therefore predicted that the *trans*-spliced EMR2 product EMR2(1-2-3*-4-5), but not EMR2(1-2-3*-5), would also bind to CS. To test this, a modified cell-capture binding assay was performed using a well-established mFc-fusion protein approach [43,44]. mFc-fusion proteins comprising the wild-type EMR2(1-2-3-4-5), CD97(1-2-3-4-5), as well as EMR2(1-2-3*-4-5) and EMR2(1-2-3*-5) were generated independently (Fig. 3A), purified, and coated onto Protein G-conjugated Dynabeads. The protein–bead complexes were subsequently mixed with various CHO-K1 cell lines. Following careful washing, bound cells were eluted in EDTA solution and counted. As predicted, EMR2(1-2-3*-4-5) but not EMR2(1-2-3*-5) proteins bound to the wild-type CHO-K1 cells and even better to CHO-Pgs677 cells, a mutant CHO cell

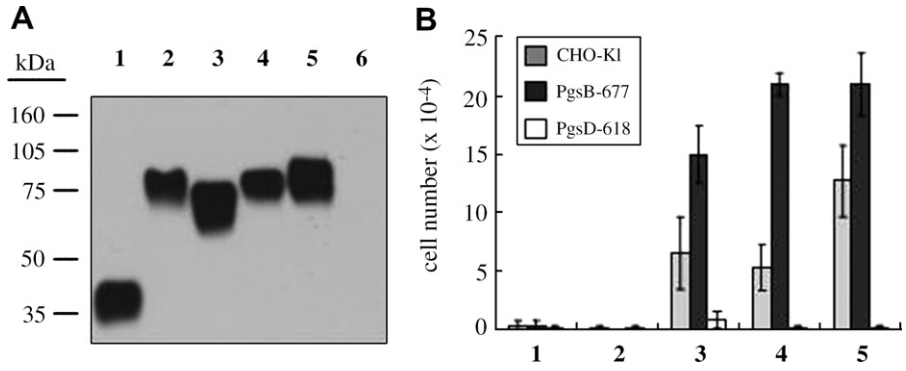

Fig. 3. Ligand-binding assay of alternatively spliced EMR2 isoforms. (A) Western blot analysis of purified mFc fusion proteins. An Fc specific anti-mouse IgG-HRP was used to detect the mFc-fusion proteins: mFc alone (lane 1), EMR2(1-2-3-4-5)-mFc (lane 2), EMR2(1-2-3*-5)-mFc (lane 3), EMR2(1-2-3*-4-5)-mFc (lane 4), CD97(1-2-3-4-5)-mFc (lane 5), and medium control (lane 6). (B) The ligand-binding assay showing the number of different CHO-K1 cell lines captured by the Dynabead complexes containing mFc alone (lane 1), EMR2(1-2-3*-5)-mFc (lane 2), CD97(1-2-3-4-5)-mFc (lane 3), EMR2(1-2-3-4-5)-mFc (lane 4), or EMR2(1-2-3*-4-5)-mFc (lane 5). The wild-type CHO-K1 cells, PgsB-677, and PgsD-618 mutant cells analyzed were represented by grey, black, and white bars, respectively. Data shown were means ± S.D. of three independent experiments.

line defective in heparan sulfate (HS)-GAG synthesis but with a higher level of cell surface CS-GAG (Fig. 3B) [43]. The cell-binding by EMR2(1-2-3*-4-5) is comparable to the wild-type EMR2(1-2-3-4-5) and CD97(1-2-3-4-5) probes, consistent with the previous finding that the CS-binding activity is mediated by the 4th EGF-like motif [43]. All probes fail to bind CHO-Pgs618 cells, another mutant CHO cell line that is deficient in all types of GAG biosynthesis (Fig. 3B) [43], confirming the binding specificity. These results indicate that the products of the *trans*-spliced EMR2 transcripts are functional in ligand-binding. Whether the *trans*-spliced EMR2 chimeric proteins are capable of binding to other novel cellular ligand(s) is currently unknown and is of great interest.

3.4. Potential mechanism and implication of alternative trans-splicing in the EGF-TM7 protein diversity

Although the molecular mechanisms involved in mediating and regulating alternative *trans*-splicing in eukaryotes are poorly understood, recent studies have slowly unveiled factors that might be relevant in these processes [5,8]. In addition to the numerous protein components also employed in *cis*-splicing, it was reported that certain sequence elements could potentially promote *trans*-splicing. These include the purine-rich exon-recognition sequence (ERS, or exonic-splicing enhancer, ESE) in the downstream exon [19,47,48], the 5′ cryptic splice sites found in some viral RNAs [15,18], and the presence of a long intron and/or of an RNA polymerase II pause site within an intron [49]. As for the regulation of *trans*-splicing, nerve growth factor has been shown to up-regulate the *trans*-splicing of a voltage-gated sodium channel in rat neurons [21].

A careful examination of the EGF-3 motif of CD97 and EMR2 has identified 3 potential purine-rich ERS motifs (Fig. 2A). Whether they play a role in EMR2/CD97 *trans*-splicing will require future study. In addition, we have also compared the length of the introns in the region containing the EGF-like motifs. It is interesting to note that in both EMR2 and CD97 the size of the 4th intron between the EGF-2 and EGF-3 exons has all exceeded more than 5 kbp (values are italicized), a much longer sequence in comparison to other introns (Table 2). Takahara et al. had demonstrated that delay in the 3′ splice site synthesis caused by the preceding intron as short as 4.1 kbp may trigger intragenic *trans*-splicing [49]. It is tempting to suspect that the long 4th intron might also contribute to the intergenic *trans*-splicing of the CD97 EGF-3* motif identified here.

Table 2
The exon–intron organization of EMR2 gene

Exon	Exon size (bp)		Intron	Intron size (kbp)	
	EMR2	CD97		EMR2	CD97
1 (5′-UTR + SP)[a]	⩾101	⩾145	1	2.42	6.9
2 (SP)	51	51	2	0.20	0.20
3 (EGF-1)	117	117	3	1.44	2.10
4 (EGF-2)	156	156	4	*5.23*	*5.26*
5 (EGF-3)	132	132	5	0.60	0.60
6 (EGF-4)	147	147	6	0.43	0.43
7 (EGF-5)	147	147	7	0.09	0.09

[a]UTR and SP stand for untranslated region and signal peptide, respectively.

As EMR2 and CD97 are believed to be duplicated from an ancestor gene and share strong homology, we wonder whether the reciprocal *trans*-splicing event would also occur in CD97, i.e. CD97 transcripts containing the EGF-like 3 motif of EMR2. Surprisingly, a search in the EST databases to date (as of October 2007) has failed to identify any EST clone that matches to the CD97(1-2-3#-4-5) or CD97(1-2-3#-5) chimeric transcript (# denotes the sequences of EMR2 EGF-like motif) (data not shown). This result not only has strengthened the authenticity of the *trans*-spliced EMR2 sequences reported here, but has also further pointed to the unique characteristic of the EMR2 chimeric transcripts, suggesting an underlying functional role for the chimeric proteins they translated. The differences in the *trans*-splicing capacity of EMR2 and CD97 genes might provide an opportunity to further dissect the molecular mechanisms involved. For example, the expression levels of EMR2 and CD97 in different myeloid cell types have been found to be variable, which might be a factor in promoting *trans*-splicing [42,50].

In conclusion, we have presented strong evidence that alternative *trans*-splicing occurs between EMR2 and CD97 genes. The resulting EMR2 chimeric proteins contain a specific EGF-like domain (EGF-3) from a closely related gene, CD97, and are functional in ligand-binding. As well as alternative

cis-splicing, we suggest that *trans*-splicing is a novel mechanism for the EGF-TM7 gene family to further expand its protein repertoire.

Acknowledgements: The authors would like to thank Drs. M. Stacey and J.Q. Davies for the critical reading of the manuscript. This study was supported by research grants from Chang Gung Memorial Hospital (CMRPD330083 and CMRPD160061) and the National Science Council (NSC96-2320-B-182-005) to H.-H. Lin. S. Gordon is supported by the British Heart Foundation and Medical Research Council.

References

[1] Pennisi, E. (2000) Human genome project. And the gene number is? Science 288, 1146–1147.
[2] Cheng, J. et al. (2005) Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. Science 308, 1149–1154.
[3] Black, D.L. (2000) Protein diversity from alternative splicing: a challenge for bioinformatics and post-genome biology. Cell 103, 367–370.
[4] Graveley, B.R. (2001) Alternative splicing: increasing diversity in the proteomic world. Trends Genet. 17, 100–107.
[5] Maniatis, T. and Tasic, B. (2002) Alternative pre-mRNA splicing and proteome expansion in metazoans. Nature 418, 236–243.
[6] Modrek, B. and Lee, C. (2002) A genomic view of alternative splicing. Nat. Genet. 30, 13–19.
[7] Sorek, R. and Amitai, M. (2001) Piecing together the significance of splicing. Nat. Biotechnol. 19, 196.
[8] Horiuchi, T. and Aigaki, T. (2006) Alternative trans-splicing: a novel mode of pre-mRNA processing. Biol. Cell 98, 135–140.
[9] Nilsen, T.W. (2001) Evolutionary origin of SL-addition trans-splicing: still an enigma. Trends Genet. 17, 678–680.
[10] Blumenthal, T. (1995) Trans-splicing and polycistronic transcription in *Caenorhabditis elegans*. Trends Genet. 11, 132–136.
[11] Dorn, R. and Krauss, V. (2003) The modifier of mdg4 locus in Drosophila: functional complexity is resolved by trans splicing. Genetica 117, 165–177.
[12] Dorn, R., Reuter, G. and Loewendorf, A. (2001) Transgene analysis proves mRNA trans-splicing at the complex mod(mdg4) locus in Drosophila. Proc. Natl. Acad. Sci. USA 98, 9724–9729.
[13] Goeke, S., Greene, E.A., Grant, P.K., Gates, M.A., Crowner, D., Aigaki, T. and Giniger, E. (2003) Alternative splicing of lola generates 19 transcription factors controlling axon guidance in Drosophila. Nat. Neurosci. 6, 917–924.
[14] Horiuchi, T., Giniger, E. and Aigaki, T. (2003) Alternative trans-splicing of constant and variable exons of a Drosophila axon guidance gene, *lola*. Genes Dev. 17, 2496–2501.
[15] Eul, J., Graessmann, M. and Graessmann, A. (1995) Experimental evidence for RNA trans-splicing in mammalian cells. Embo J. 14, 3226–3235.
[16] Konarska, M.M., Padgett, R.A. and Sharp, P.A. (1985) Trans splicing of mRNA precursors in vitro. Cell 42, 165–171.
[17] Solnick, D. (1985) Trans splicing of mRNA precursors. Cell 42, 157–164.
[18] Caudevilla, C., Da Silva-Azevedo, L., Berg, B., Guhl, E., Graessmann, M. and Graessmann, A. (2001) Heterologous HIV-nef mRNA trans-splicing: a new principle how mammalian cells generate hybrid mRNA and protein molecules. FEBS Lett. 507, 269–279.
[19] Caudevilla, C., Serra, D., Miliar, A., Codony, C., Asins, G., Bach, M. and Hegardt, F.G. (1998) Natural trans-splicing in carnitine octanoyltransferase pre-mRNAs in rat liver. Proc. Natl. Acad. Sci. USA 95, 12185–12190.
[20] Frantz, S.A., Thiara, A.S., Lodwick, D., Ng, L.L., Eperon, I.C. and Samani, N.J. (1999) Exon repetition in mRNA. Proc. Natl. Acad. Sci. USA 96, 5400–5405.
[21] Akopian, A.N., Okuse, K., Souslova, V., England, S., Ogata, N. and Wood, J.N. (1999) Trans-splicing of a voltage-gated sodium channel is regulated by nerve growth factor. FEBS Lett. 445, 177–182.
[22] Flouriot, G., Brand, H., Seraphin, B. and Gannon, F. (2002) Natural trans-spliced mRNAs are generated from the human estrogen receptor-alpha (hER alpha) gene. J. Biol. Chem. 277, 26244–26251.
[23] Finta, C. and Zaphiropoulos, P.G. (2002) Intergenic mRNA molecules resulting from trans-splicing. J. Biol. Chem. 277, 5882–5890.
[24] Li, B.L. et al. (1999) Human acyl-CoA:cholesterol acyltransferase-1 (ACAT-1) gene organization and evidence that the 4.3-kilobase ACAT-1 mRNA is produced from two different chromosomes. J. Biol. Chem. 274, 11060–11071.
[25] Tasic, B. et al. (2002) Promoter choice determines splice site selection in protocadherin alpha and gamma pre-mRNA splicing. Mol. Cell 10, 21–33.
[26] Fitzgerald, C., Sikora, C., Lawson, V., Dong, K., Cheng, M., Oko, R. and van der Hoorn, F.A. (2006) Mammalian transcription in support of hybrid mRNA and protein synthesis in testis and lung. J. Biol. Chem. 281, 38172–38180.
[27] Coady, T.H., Shababi, M., Tullis, G.E. and Lorson, C.L. (2007) Restoration of SMN function: delivery of a trans-splicing RNA re-directs SMN2 pre-mRNA splicing. Mol. Ther. 15, 1471–1478.
[28] Wood, M., Yin, H. and McClorey, G. (2007) Modulating the expression of disease genes with RNA-based therapy. PLoS Genet. 3, e109.
[29] McKnight, A.J. and Gordon, S. (1996) EGF-TM7: a novel subfamily of seven-transmembrane-region leukocyte cell-surface molecules. Immunol. Today 17, 283–287.
[30] McKnight, A.J. and Gordon, S. (1998) The EGF-TM7 family: unusual structures at the leukocyte surface. J. Leukoc. Biol. 63, 271–280.
[31] Fredriksson, R., Gloriam, D.E., Hoglund, P.J., Lagerstrom, M.C. and Schioth, H.B. (2003) There exist at least 30 human G-protein-coupled receptors with long Ser/Thr-rich N-termini. Biochem. Biophys. Res. Commun. 301, 725–734.
[32] Stacey, M., Lin, H.H., Gordon, S. and McKnight, A.J. (2000) LNB-TM7, a group of seven-transmembrane proteins related to family-B G-protein-coupled receptors. Trends Biochem. Sci. 25, 284–289.
[33] Kwakkenbos, M.J., Kop, E.N., Stacey, M., Matmati, M., Gordon, S., Lin, H.H. and Hamann, J. (2004) The EGF-TM7 family: a postgenomic view. Immunogenetics 55, 655–666.
[34] Hamann, J., Hartmann, E. and van Lier, R.A. (1996) Structure of the human CD97 gene: exon shuffling has generated a new type of seven-span transmembrane molecule related to the secretin receptor superfamily. Genomics 32, 144–147.
[35] Kwakkenbos, M.J. et al. (2006) An unusual mode of concerted evolution of the EGF-TM7 receptor chimera EMR2. Faseb J. 20, 2582–2584.
[36] Lin, H.H., Stacey, M., Hamann, J., Gordon, S. and McKnight, A.J. (2000) Human EMR2, a novel EGF-TM7 molecule on chromosome 19p13.1, is closely related to CD97. Genomics 67, 188–200.
[37] Gray, J.X. et al. (1996) CD97 is a processed, seven-transmembrane, heterodimeric receptor associated with inflammation. J. Immunol. 157, 5438–5447.
[38] Hamann, J. et al. (1995) Expression cloning and chromosomal mapping of the leukocyte activation antigen CD97, a new seven-span transmembrane molecule of the secretion receptor superfamily with an unusual extracellular domain. J. Immunol. 155, 1942–1950.
[39] Hamann, J., Stortelers, C., Kiss-Toth, E., Vogel, B., Eichler, W. and van Lier, R.A. (1998) Characterization of the CD55 (DAF)-binding site on the seven-span transmembrane receptor CD97. Eur. J. Immunol. 28, 1701–1707.
[40] Hamann, J., Vogel, B., van Schijndel, G.M. and van Lier, R.A. (1996) The seven-span transmembrane receptor CD97 has a cellular ligand (CD55, DAF). J. Exp. Med. 184, 1185–1189.
[41] Lin, H.H. et al. (2001) Molecular analysis of the epidermal growth factor-like short consensus repeat domain-mediated protein–protein interactions: dissection of the CD97–CD55 complex. J. Biol. Chem. 276, 24160–24169.
[42] Kwakkenbos, M.J., Pouwels, W., Matmati, M., Stacey, M., Lin, H.H., Gordon, S., van Lier, R.A. and Hamann, J. (2005) Expression of the largest CD97 and EMR2 isoforms on leukocytes facilitates a specific interaction with chondroitin sulfate on B cells. J. Leukoc. Biol. 77, 112–119.

[43] Stacey, M., Chang, G.W., Davies, J.Q., Kwakkenbos, M.J., Sanderson, R.D., Hamann, J., Gordon, S. and Lin, H.H. (2003) The epidermal growth factor-like domains of the human EMR2 receptor mediate cell attachment through chondroitin sulfate glycosaminoglycans. Blood 102, 2916–2924.
[44] Lin, H.H., Stacey, M., Chang, G.W., Davies, J.Q. and Gordon, S. (2005) Method for selecting and enriching cells expressing low affinity ligands for cell surface receptors. Biotechniques 38, 696–698.
[45] Akiva, P., Toporik, A., Edelheit, S., Peretz, Y., Diber, A., Shemesh, R., Novik, A. and Sorek, R. (2006) Transcription-mediated gene fusion in the human genome. Genome Res. 16, 30–36.
[46] Parra, G., Reymond, A., Dabbouseh, N., Dermitzakis, E.T., Castelo, R., Thomson, T.M., Antonarakis, S.E. and Guigo, R. (2006) Tandem chimerism as a means to increase protein complexity in the human genome. Genome Res. 16, 37–44.
[47] Caudevilla, C. et al. (2001) Localization of an exonic splicing enhancer responsible for mammalian natural trans-splicing. Nucleic Acids Res. 29, 3108–3115.
[48] Tanaka, K., Watakabe, A. and Shimura, Y. (1994) Polypurine sequences within a downstream exon function as a splicing enhancer. Mol. Cell Biol. 14, 1347–1354.
[49] Takahara, T., Tasic, B., Maniatis, T., Akanuma, H. and Yanagisawa, S. (2005) Delay in synthesis of the 3′ splice site promotes trans-splicing of the preceding 5′ splice site. Mol. Cell 18, 245–251.
[50] Hamann, J. et al. (2007) EMR1, the human homolog of F4/80, is an eosinophil-specific receptor. Eur. J. Immunol. 37, 2797–2802.